DOCUMENT RESUME

ED 290 761 TM 011 001

AUTHOR Stiggins, Richard J.; And Others
TITLE Measuring Thinking Skills through Classroom Assessment.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Center for Performance Assessment.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 87
CONTRACT 400-86-0006
NOTE 44p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Techniques; *Critical Thinking; Elementary Secondary Education; Evaluation Methods; Inferences; Language Arts; Mathematics Skills; *Measurement Techniques; *Recall (Psychology); Science Education; Social Studies; *Student Evaluation
IDENTIFIERS *Higher Order Skills; *Thinking Skills

ABSTRACT

Classroom assessment procedures of 36 teachers in grades 2 to 12 from the same suburban school district in the Pacific Northwest were studied to determine the extent to which they measure higher order thinking skills of students in mathematics, science, social studies, and language arts. Gathering information from teachers involved four steps: (1) teachers were interviewed about their plans for one instructional day; (2) teachers were each observed by a trained observer for all class periods of the day; (3) teachers provided observers with four to six samples of paper and pencil assessment instruments used recently; and (4) each was interviewed indepth after the day of observation. In both written and oral assessments, teachers focused nearly half of their assessments on the simple recall of facts and information. Inference and analysis received less attention, and comparison and evaluation were almost ignored. Despite teachers' understanding of the importance of teaching students to think, they still require students to reproduce facts and information. Only mathematics emphasized thinking beyond mere recall. Teachers at various grade levels were surprisingly similar in their assessment patterns. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

THE Northwest Regional Educational Laboratory

PROGRAM REPORT

ED290761

MEASURING THINKING SKILLS THROUGH CLASSROOM ASSESSMENT

Richard J. Stiggins
Maggie Griswold
Karen Reed Green
and
Associates

November 1987

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JERRY D. KIRKPATRICK

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM011001

Sponsored by

OERI Office of Educational Research and Improvement
U.S. Department of Education

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number 400-86-0006. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

MEASURING THINKING SKILLS

THROUGH CLASSROOM ASSESSMENT

Richard J. Stiggins, Maggie Griswold,
Karen Reed Green and Associates*

November 1987

Center for Performance Assessment
Northwest Regional Educational Laboratory
101 S.W. Main, Suite 500
Portland, Oregon 97204

* Others associated with the successful completion of this study include the field research team members Marilyn Hartzell, Lynde Paule, Karen Nelson and Nelson Doelman. Special thanks to Jim Carlile and to the 36 who opened their classrooms and gave their time.

Abstract

The classroom assessment procedures of 36 teachers in grades 2 to 12 were studied in depth to determine the extent to which they measure students' higher order thinking skills in mathematics, science, social studies and language arts. A wide variety of assessment documents were analyzed, teachers were observed in their classrooms and each teacher was interviewed. The results suggest that paper and pencil assessments are dominated by assessment of information by recall across all grade levels. However, inference is assessed also, especially in mathematics. Oral assessments employ mostly recall questions, with analysis and inference reflected in many questions. Across assessment forms, grades and subjects, comparison and evaluation questions are ignored. While teachers are trained to teach thinking skills, they are far less often trained to assess such skills. Implications of the results are explored.

MEASURING THINKING SKILLS THROUGH CLASSROOM ASSESSMENT

The assessments that contribute most to student learning and to students' academic self-concept are those that teachers use in the classroom on a day-to-day basis. A small but growing body of research is beginning to provide a clearer picture of the nature, role, quality and impact of these classroom assessments (Stiggins, Conklin, & Bridgeford, 1986; and, Natriello, 1987). Relative to large-scale assessment programs, such as district and state-wide standardized testing programs which command most of the test score publicity and assessment resources, teacher-developed and text-embedded assessments command almost all of the students' and teachers' attention in the classroom.

Because these classroom measures of student achievement are so important to the learning process, we need to be certain that they are of very high quality. Unfortunately, however, there is reason to believe that they are not as sound as they might be. For instance, Stiggins and Bridgeford (1985) report quality control problems with classroom assessments based on teacher observation and judgment, and Fleming and Chambers (1983) report problems with the quality of teacher-developed paper and pencil tests. Further, Schafer and Lisstz (1987) have documented in very clear terms the inadequacy of the measurement training programs offered to educators.

Couched within this general issue of classroom assessment quality, there is a more specific problem that arises from the high educational priority currently being placed on the development of higher order thinking skills. Results of national, state and local testing programs in recent years have documented deficiencies in students' problem solving and critical thinking abilities. This has precipitated major curriculum development efforts in

nearly every state department of education, through many professional associations and in many local school districts. These developmental efforts have included a great deal of excellent conceptual thought and planning (see Baron & Sternberg, 1985, for example) and have given rise to the addition of thinking skills assessments to state-wide tests and published standardized tests (Kearney, Kean, Raeber, Stevens, Baron, Fremer & Daniel, 1985).

But once again, as with other important advances in assessment in the past, there is the danger that all of the research, conceptual work and implementation will take place in the context of the highly-visible large-scale testing programs and inadequate attention and resources will be given to documenting and improving the quality of classroom assessments of those same skills. If we are to succeed in achieving instructional objectives related to the development of higher order thinking skills, we need to (a) plan and present quality educational experiences that develop those skills and (b) use classroom assessments that match the objectives and instructional experiences provided to students. It does little good, for example, to aspire to the development of problem solving skills, attempt to teach those skills and then document the students' achievement of such lofty objectives by measuring whether they can memorize and recall simple facts.

Unfortunately, there is reason to believe that this is exactly what is happening. In a study of the test development practices of nearly 200 high school teachers, Reynolds and Menard (1980) analyzed teachers' self-report data on their item writing practices and found a very heavy reliance on items written at the knowledge, comprehension and applications levels of Bloom's taxonomy. Very little attention was paid, by contrast, to analysis, synthesis and evaluation levels.

In an even more focused study, Fleming and Chambers (1983) went far beyond the self-report data used in the previously cited study to examine actual test items taken from teacher-developed tests. The results provide a graphic picture of extent to which higher order thinking skills are assessed in the classrooms of one urban school district:

Grade Level	Know. Level	Comp. Level	Analy. Level	App. Level	Total Items
Elementary	83%		10%	7%	3176
Junior High	97%		3%		3483
High School	88%	9%	3%		2160
				Total	8819

These items covered a wide variety of subject matter areas, including language arts, math, social studies, science, industrial arts, and French. Thus, there is reason to be concerned about assessments of thinking skills in these classrooms.

In yet another relevant study, Carter (1984) examined teachers' level of comfort with and actual ability to write test items at different levels of thinking skills. Teachers had little difficulty writing items at the recall level but had great difficulty and experienced considerable discomfort working at the higher levels of Bloom's taxonomy.

The limited scope of these initial studies restricts the generalizability of results. For instance, the Reynolds and Menard study tapped teachers' perceptions of their test development procedures, but did not follow up to determine if perceptions matched actual practice. Further, the Fleming and Chambers research ignored teachers' intentions, thus also precluding any

attempt to compare intentions with actual tests. In addition, they focused only on teacher-developed paper and pencil tests, which represent only a fraction of the assessment devices used by teachers in classrooms. Finally, it is not clear whether Fleming and Chambers examined test items only in their analysis or if they also examined the text material which the tests were designed to cover. Examination of tests and text is essential in order to classify an item as to cognitive level.

Because of these shortcomings in previous research and the need to develop an expanded data base on classroom assessment of thinking skills, we designed an expanded replication of the Fleming and Chambers study. First, like the previous investigators, we sought to explore variations in the extent to which the assessments of thinking skills varied across a wide range of grade levels and a variety of school subjects. In addition, however, we intended to examine five different types of assessment tools, all of which make major contributions to classroom assessment environments: teacher-developed paper and pencil tests, text-embedded tests, written assignments, performance assessments, and oral questions posed of students during instruction. Each of these vehicles is capable of measuring thinking skills. The question is which are used in this way by teachers?

Finally, the data collection procedures were designed to allow a comparision of the teachers' assessment objectives and the actual assessments they develop or select to serve those objectives.

As mentioned above, the validity of such studies rests on the accuracy of the classification of each question with respect to its cognitive level. Questions can only be properly classified when considered in relation to the material they are intended to test. For example, questions that appear to

test analysis or inference may only require regurgitation of material covered explicitly in the text. One can only know this by comparing questions to text. Such comparisons formed the basis for the coding in this study.

METHODOLOGY

The procedures for gathering information from teachers regarding their assessments included four steps. First, teachers were interviewed briefly to gather data on their plans for one instructional day; then each teacher was observed by a trained observer for all class periods during that day; third, the teachers provided the investigator with between four and six samples of paper and pencil assessment instruments used recently; and finally, each was interviewed indepth shortly after the day of observation.

Teachers Studied

The assessment instruments and procedures of 36 volunteer teachers were studied. They were distributed equally across the following grade level categories: 1-2, 3-4, 5-6, 7-8, 9-10, and 11-12, with emphasis on the even grades. All taught in one high school, one middle school, or one of three elementary schools in the same suburban district in the Pacific Northwest. All schools and the district had set major long-term goals to teach thinking skills and some inservice training had already begun. However, specifics of school or district programs had not yet been drawn together.

For this study we focused on four core subjects: mathematics, science, social studies, and language arts. Table 1 presents the distribution of teachers across grades and subject areas. The elementary teachers generally taught a cross-section of subjects. The junior high and high school teachers,

Table 1
Distribution of Teachers, Assessment Documents and Oral Questions
Across Grades and Subjects

		Subject Area				
Grade	Element	Mathematics	Science	Social Studies	Language Arts	Total
1-2	# of teachers teaching	6	3	3	6	18*
	# of documents analyzed	12	3	2	17	35**
	# of oral questions analyzed	154	31	67	483	735
3-4	# of teachers teaching	5	4	4	5	18*
	# of documents analyzed	6	5	5	7	23
	# of oral questions analyzed	161	166	113	423	863
5-6	# of teachers teaching	5	4	3	6	18*
	# of documents analyzed	3	5	2	10	20
	# of oral questions analyzed	201	109	5	407	722
7-8	# of teachers teaching	1	1	2	2	6*
	# of documents analyzed	6	7	3	7	23
	# of oral questions analyzed	199	218	181	422	1020
9-10	# of teachers teaching	2	2	1	5	10*
	# of documents analyzed	5	6	0	11	22
	# of oral questions analyzed	64	133	66	299	562
11-12	# of teachers teaching	2	2	1	4	9*
	# of documents analyzed	7	3	9	7	26
	# of oral questions analyzed	92	97	440	211	840
Total	# of teachers teaching	21	16	14	28	79*
	# of documents analyzed	39	29	21	59	149**
	# of oral questions analyzed	871	754	872	2245	4742

* Denotes teacher/subject combinations, rather then total number of teachers in the study.

** One extra document is included for which subject identification was missing.

however, were more specialized, although some taught multiple subjects. This teaching of several subjects is reflected in the totals of teachers by subject by grade level, which denote teacher/subject combinations and therefore total far more than the 36 teachers in the study. As might be expected, there were more teacher/subject combinations in grades 1-6.

Data Collection Methods

All data on teachers' assessment procedures were collected in the spring of the 1986-87 academic year. In preparation, data collection forms were developed and pilot tested by the field staff. After the field trial, forms and procedures were revised. Since the entire team of five field staff and the principal investigator worked cooperatively to design, field test, and refine the data collection forms and procedures, preparation for the research served as the training ground for field staff.

All teachers were recruited by their principal to participate. They were called to a meeting in their school, where a member of the research team explained the rationale for the study and the commitments participants would be making if they agreed to be part of the study. It was described as a study of classroom assessment procedures in general. No explicit mention was made of the focus on higher order thinking skills, so as not to bias the study. Teachers were advised that participation was voluntary. Only one teacher dropped out of the sample.

Initial interview. Each member of the field staff was assigned a group of teachers to study. The staff's first step was to set up an initial meeting with each teacher to review the purpose of the study, explain the data that would be collected, select a day for classroom observation, and review in detail the teacher's instructional plans for that day. The data collection

form (see appendix) provided space for the researcher to record for each class period during the day: background information, including the course to be taught, number of students, grade, an estimate of the ability of the students in that class, and class meetings per week; instructional information, including goals, new or review material to be covered, and activities planned; and, any assessment plans for that period. The researcher also obtained copies of texts and other materials and information specifically on the material to be covered that day.

Classroom observation. The researcher then returned on the appointed day to observe for the entire day or as much of it as was feasible. During each class period, the observer concentrated on the questions posed by the teacher. Each question was coded as to the level of thinking it reflected (see Thinking Skills Framework section below), the respondent (targeted or volunteer), and the correctness of the response. The pilot test of observational procedures with multiple observers in the same classroom coding the same questions verified the objectivity of the coding system.

A total of 4,742 oral questions were analyzed, from all of the participating teachers and across the four subject areas. As Table 1 indicates, nearly half of these questions were posed in language arts classes. This is due in part to the fact that approximately one-third of the classes (teacher/subject combinations) we observed were language arts. However, more oral questions were asked and coded in language arts classes than in the other subject areas also.

Document analysis. On the observation day, the researcher also secured from the teacher four to six samples of paper and pencil assessments used recently in that classroom. This included any used on observation day plus others of the teacher's choosing. These written assessments were then

analyzed and coded as to: their nature, including teacher-developed paper and pencil test, text-embedded test, written assignment, performance assessment, standardized test, or other type; subject matter covered, author, and use; the number and proportion of items that reflected each level of cognitive operation; and, the number and percent of items of different formats, such as selection, fill-in, essay, product, and other. All document analyses were conducted first by the field researcher responsible for that teacher and then were reviewed by a second research team-member to assure proper coding.

A total of 149 assessment documents were analyzed. These were distributed evenly across grade level, with the exception of a larger number at grades 1-2. Distribution across subjects was more uneven. The fewest were collected in social studies, the most in language arts. As with the oral questions, this is partially due to the fact that we observed more language arts classes.

Performance assessments, which are measures based on teacher observation and judgment, were coded in more detail. These records reflected the nature of the performance observed and judged by the teacher, including process and products, as well as the clarity of the performance criteria, and the levels of thinking skills reflected in the criteria. Performance assessments are not reported here because we observed so few.

Final interview. After the documents had been coded and the observation records reviewed, the field staff member returned to the school for a final interview with the teacher. The interview protocol called for the researcher to probe the teacher's use of and attitudes about six different types of classroom assessment procedures: teacher-developed paper and pencil tests and quizzes, text-embedded paper and pencil tests and quizzes, written

assignments, performance assessments, oral questions, and standardized achievement tests. Teachers were asked to describe their use of each of these to measure higher order thinking skills. In addition, they were queried about their training in thinking skills assessment and instruction, and their attitudes about instruction in this arena.

Data collection forms and instructions for both interviews, the classroom observations, and the document analyses are appended to this report.

The Thinking Skills Framework

In order to generate data on the classroom assessment of higher order thinking skills that are comparable across teachers, grades, school subjects, and other important independent variables, it was necessary for the study to focus on one framework or taxonomy of thinking skills. Of the many available options, we elected to apply the framework outlined by Quellmalz (1985). Five levels of cognitive operations are included: recall, analysis, comparison, inference, and evaluation. Each level is described in detail below, along with its relation to the more commonly used Bloom taxonomy.

This particular framework was selected for a number of reasons. First and most importantly, Quellmalz makes a compelling argument that these five levels collect all of the elements common to a great many other taxonomic structures of thinking skills. Second, the levels included in the Quellmalz structure are conceptually simple, making coding of questions relatively easy. Third, the field staff had developed teacher training programs centered on this taxonomy and therefore were very familiar with it. This minimized the need for field staff training.

Chart 1
SUMMARY OF THINKING SKILLS

Quellmalz Level	Definition	Relation to Bloom Taxonomy
Recall	Most tasks require that students recognize or remember key facts, definitions, concepts, rules, and principles. Recall questions require students to repeat verbatim or to paraphrase given information. To recall information, students need most often to rehearse or practice it, and then to associate it with other, related concepts. The Bloom taxonomy levels of knowledge and comprehension are subsumed here, since verbatim repetition and translation into the student's own words represent acceptable evidence of learning and understanding.	Knowledge Comprehension
Analysis	In this operation, students divide a whole into component elements. Generally the part/whole relationships and the cause/effect relationships that characterize knowledge within subject domains are essential components of more complex tasks. The components can be the distinctive characteristics of objects or ideas, or the basic actions of procedures or events. This definition of analysis is the same as that in the Bloom taxonomy.	Analysis
Comparison	These tasks require students to recognize or explain similarities and differences. Simple comparisons require attention to one or a few very obvious attributes or component processes, while complex comparisons require identification of the differentiation among many attributes or component actions. This category relates to some of the skills in the Bloom level of analysis. The separate comparison category emphasizes the distinct information processing required when students go beyond breaking the whole into parts in order to compare similarities and differences.	Analysis
Inference	Both deductive and inductive reasoning fall into this category. In deductive tasks, students are given a generalization and are required to recognize or explain the evidence that relates to it. Applications of rules and "if then" relationships require inference. In inductive tasks, students are given the evidence or details and are required to come up with the generalization. Hypothesizing, predicting, concluding, and synthesizing all require students to relate and integrate information. Inductive and deductive reasoning relate to the Bloom levels of application and synthesis. Application of a rule is one kind of deductive reasoning, synthesis, putting parts together to form a generalization, occurs in both inductive and deductive reasoning.	Application Synthesis
Evaluation	These tasks require students to judge quality, credibility, worth, or practicality. Generally we expect students to use established criteria and explain how these criteria are or are not met. The criteria might be established rules of evidence, logic, or shared values. Bloom's levels of synthesis and evaluation are involved in this category. To evaluate, students must assemble and explain the interrelationship of evidence and reasons in support of their conclusion (synthesis). Explanation of criteria for reaching a conclusion is unique to evaluative reasoning.	Synthesis

5530e

RESULTS

The teachers in this study use a variety of types of assessments in the classroom. They also believe that higher order thinking skills are very important in student learning. However, they tend to limit their assessments of student thinking to only one or two levels: recall and inference. The other levels of thinking--analysis, comparison and evaluation--are assessed much less frequently.

The vast majority of these teachers have been trained to teach higher order thinking skills. Three-quarters of them have participated in more than one training workshop on the topic. However, fewer than one-third of them have had more than one training session in the assessment of higher order skills, and over a third have not had any training at all in this area. This may explain the finding that they tend not to assess the full range of thinking skills.

In the pages that follow, we present the results of this study in greater detail. We then discuss these findings, considering possible interpretations and implications for teacher training and for further research.

Assessment Document Analysis

From the 36 teachers we collected and analyzed 149 assessment documents. As Table 2 indicates, these included paper and pencil tests or quizzes developed by the teacher, paper and pencil tests or quizzes taken from the text being used for the course, written assignments, standardized tests, and other types of written assessments.

The largest number of documents analyzed were paper and pencil tests and quizzes developed by the teacher. These comprised 38% of the documents gathered. Another 30% were tests and quizzes that accompany the textbooks.

Written assignments (study questions, compositions and the like) made up another 20%. Many of these assignments also were taken from the texts. Thus, based on the documents teachers provided us, there appears to be a heavier reliance on text-embedded assessments than indicated in Table 2.

Table 2
Type of Documents Analyzed

Type	N	Percent
Teacher-Developed Paper and Pencil Test	57	38
Text-Embedded Paper and Pencil Test	45	30
Written Assignment	30	20
Standardized Test	7	5
Other	10	7
TOTAL	149	100

If we look at the distribution of these various types of assessment documents by grade level (Table 3), we see a somewhat heavier reliance on text-embedded tests and quizzes in the lower grades. In middle school and high school, teachers tend to create their own assessments to a much greater degree, according to these data.

Table 3
Distribution of Assessment Documents Analyzed by Grade and Type of Document, in Percent

	Type of Document					
Grade	Tea-Dev P&P Test	Text P&P Test	Written Assign.	Standardized Achievement Test	Other	Total All Types
1-2	11	63	0	6	20	100
3-4	30	30	17	13	9	99
5-6	35	45	15	5	0	100
7-8	65	0	35	0	0	100
9-10	36	23	32	5	5	101
11-12	62	8	31	0	0	101
Total all grades	38	30	20	5	7	100

Analysis of the items in each document by the level of higher order thinking they assess reveals some interesting patterns. Table 4 presents this analysis by level of thinking skill for the grades taught by the participating teachers and for the four subject areas selected for this study, summarized across all grades.

Table 4
Percent of Document Items Assessing Higher Order Thinking Skills
by Level Tested, by Grade and by Subject Area Across Grades

	Level of High Order Thinking				
Grade	Recall %	Analysis %	Comparison %	Inference %	Evaluation %
1-2	56	12	14	19	0
3-4	41	16	4	34	5
5-6	44	19	4	30	3
7-8	51	7	1	39	2
9-10	42	12	3	39	5
11-12	41	9	4	44	3
Total	46	12	6	33	3
Subject (across grades)					
Math	19	0	9	72	0
Science	65	11	5	17	2
Social Studies	66	14	5	13	3
Language Arts	49	19	5	23	5

Clearly, the largest percent of items test recall of facts and information. Nearly half of all items found in these documents assess recall. This reliance on recall is strong at all grade levels. Inference is also a level of thinking frequently assessed in these documents. Indeed, approximately 80% of the document items analyzed test either recall or inference. Items requiring analysis level thinking were much less commonly used, and comparison and evaluation were hardly assessed at all.

When we look at the distribution of items by subject matter, we find a slightly different pattern. We find a very heavy reliance on recall, but not in all subjects. In science and social studies classes these teachers make particularly strong use of recall (65 and 66%, respectively). In language arts it is also commonly used (49%). In mathematics, however, the pattern is strikingly different: Only 19% of the items assess recall, whereas 72%--nearly three-quarters--of all items tap the inference level of thinking.

The overwhelming predominance of inference level items in math must be considered when we examine the grade-level results more closely. The relatively high assessment of inference level thinking apparent at the different grade levels is heavily influenced by the preponderance of inference in math assessments. If we remove math from the grade-level analyses (see Table 5), we find an even heavier reliance on recall (55% overall, and at some grade levels over 65%), a substantial decrease in inference, and some increase in analysis level items. The paucity of comparison and evaluation items remains basically the same.

Table 5
Percent of Document Items Assessing Higher Order Thinking Skills by Level Tested, by Grade for Science, Social Studies, and Language Arts (Math Excluded)

	Level of High Order Thinking				
Grade	Recall %	Analysis %	Comparison %	Inference %	Evaluation %
1-2	66	18	7	10	0
3-4	43	20	6	20	7
5-6	45	22	5	24	4
7-8	68	10	2	18	2
9-10	53	17	3	20	7
11-12	53	12	5	23	8
Total	55	16	5	19	4

Oral Questions Analysis

From our observations of a full day of classes with each of the 36 teachers, we coded 4,742 oral questions. Table 6 displays the distribution of these questions by higher order thinking skills level for each of the grade levels observed and for each of the four subject areas, across grade level.

Table 6
Percent of Oral Questions Assessing Higher Order Thinking Skills
by Level Tested, by Grade and by Subject Area Across Grades

	Level of High Order Thinking				
Grade	Recall %	Analysis %	Comparison %	Inference %	Evaluation %
1-2	70	5	5	17	4
3-4	51	15	4	25	5
5-6	36	19	5	33	8
7-8	39	35	2	12	12
9-10	48	8	6	25	13
11-12	42	23	2	18	15
Total	47	20	4	21	9
Subject (across grades)					
Math	41	26	3	26	3
Science	59	13	8	16	5
Social Studies	48	24	2	16	10
Language Arts	45	17	3	22	13

Nearly half of the questions asked in the classrooms we observed assessed recall of facts and information. In grades 1-2 this figure was as high as 70%, while in grades 5-6 only slightly more than a third of the questions assessed recall. However, in all grades recall was the the most heavily assessed level of thinking. Questions requiring inference or analysis levels

of thinking were also fairly common (21 and 20%, respectively). Nearly 90% of all oral questions asked involved one of these three levels of thinking. Evaluation and comparison were largely ignored, especially in the elementary grades.

When we examine the distribution of questions by level of thinking by subject area, we find a similar pattern. Recall dominates, particularly in science. Analysis and inference are next most often assessed in all subjects. And evaluation and comparison are found to a much lesser extent.

Looking at the assessment of different levels of thinking skills via written document as compared to oral questions (Table 7), we see strikingly similar patterns at the grade level. Whether through written or oral assessments teachers focus nearly half their assessments on simple recall of facts and information. Inference and analysis receive considerably less attention, and comparison and evaluation are virtually ignored.

Table 7
Comparison of Higher Order Thinking Skills Assessed Via Written Documents and Via Oral Questions

Grade	Level of High Order Thinking									
	Recall %		Analysis %		Comparison %		Inference %		Evaluation %	
	Doc	Oral	Doc	Oral	Doc	Oral	Doc	Oral	Doc	Oral
1-2	56	70	12	5	14	5	19	17	0	4
3-4	41	51	16	15	4	4	34	25	5	5
5-6	44	36	19	19	4	5	30	33	3	8
7-8	51	39	7	35	1	2	39	12	2	12
9-10	42	48	12	8	3	6	39	25	5	13
11-12	41	42	9	23	4	2	44	18	3	15
Total	46	47	12	20	6	4	33	21	3	9
Subject (across grades)										
Math	19	41	0	26	9	3	72	26	0	3
Science	65	59	11	13	5	8	17	16	2	5
Social Studies	66	48	14	24	5	2	13	16	3	10
Language Arts	49	45	19	17	5	3	23	22	5	13

By subject matter--with the exception of mathematics, in both written and oral assessments recall is heavily emphasized, followed by analysis and inference. Comparison and evaluation are slighted. As noted earlier, in mathematics, on the other hand, inference is assessed much to the exclusion of other levels of thinking in the assessment documents of these teachers. However, an interesting thing apparently happens during oral questioning in math classes: The number of recall questions rises markedly, to be more than double the number found in the written documents. Analysis-level thinking assessed in oral questions rises 26%, and the focus on inference-level thinking drops from 72% to a mere 26%. We will comment on these results in the Discussion section below.

Training Experiences of Participating Teachers

As noted earlier, the school district in which this study was conducted had identified the teaching of thinking skills as a major long-term goal and some inservice training had already begun. This is reflected in the fact that the vast majority of the teachers in this study had been trained to teach higher order thinking skills (Table 8). Fewer than one in ten had received no training in this area, and three-quarters of the teachers had received more than one training workshop on how to teach critical thinking skills. This preparation was true across grade levels, with the exception of grades 1-2.

Table 8
Percent of Teachers Reporting Training Experiences in Teaching and/or Assessing Higher Order Thinking Skills, by Grade

	NUMBER OF TRAINING EXPERIENCES IN TEACHING HOTS			NUMBER OF TRAINING EXPERIENCES IN ASSESSING HOTS		
Grade	None	One	More Than One	None	One	More Than One
1-2	42%	14%	42%	57%	43%	--
3-4	--	17%	83%	67%	17%	17%
5-6	--	--	100%	17%	50%	33%
7-9	--	20%	80%	20%	40%	40%
10-12	--	27%	72%	27%	18%	55%
Total	9%	17%	74%	37%	31%	31%

Training in assessment of thinking skills, however, presents a different picture. More than a third (37%) of all the teachers in the study had received no training at all in how to assess thinking skills; approximately one-third (31%) had received only one workshop; and only one-third (31%) had received more than one. Thus, these teachers are far less prepared to assess higher order thinking skills than they are to teach them. This is especially true in the lower grades. As Table 8 indicates, training in the assessment of these skills appears to be much more common among teachers at the middle and high school levels.

DISCUSSION

Findings of Interest

The heavy reliance on assessment limited to the recall level is not a particularly striking finding. In fact, it appears that these results are very similar to those of other studies we've cited. Despite their understanding of the importance of teaching students to think and problem-solve, teachers still tend to use and create questions that merely require their students to reproduce facts and information. There are some findings, however, that we believe stimulate or deserve further comment. These are discussed here.

Suprisingly high emphasis on recall in oral questions. Because of the generally heavy workloads of most teachers, we might have expected to see a high proportion of recall questions in the written assessment documents. Teachers might perceive that this level of question requires far less time to score. However, we might have expected to see a smaller proportion of recall

questions in the oral questioning used in the classroom. It would seem that classroom discussion provides the opportunity for a teacher to guide thinking and encourage exploration of different levels of thinking. This was not the case among the majority of teachers in this study.

Startling absence of comparison and evaluation. We were also surprised to see such a strikingly low amount of assessment at the comparison and evaluation levels of thinking. Comparison is not necessarily more difficult to assess than other levels of thinking. Evaluation, perhaps, may be viewed as difficult, because there may be no "right" answer. Some teachers may not feel secure enough in the subjects they are teaching to be able to guide and in fact evaluate answers to evaluation questions. But in general the absence of comparison and evaluation in these assessments was not anticipated.

Surprising lack of variation across grade levels. The teachers at various grade levels in this study were surprisingly similar in the patterns of their assessment of higher order thinking skills. We might have expected to see increased assessment of the range of thinking skills in the higher grades. For example, one might expect more recall at the elementary level. Some teachers may feel that, especially in the early grades, teaching higher order thinking skills is not yet appropriate. Also, in grade school teachers are often juggling several different subjects and may not feel completely confident in all of them, thus feeling less willing to ask more demanding questions of their students. In middle school and especially in high school, however, we would expect to find greater emphasis on fostering critical thinking and problem-solving. Teachers tend to be more specialized in subject area, and they tend to prefer to develop their own assessment instruments.

Under these circumstances, we might expect to see more evidence of the use of the range of thinking skills in their assessments. The lack of some evolution in this dimension of assessment as one progresses through the higher grades is a topic worthy of further exploration.

Mathematics as an unusual case. At first glance one might assume it natural that mathematics was so strikingly different in its use of levels of thinking beyond recall. After all, it is problem-solving. However, the other subjects also lend themselves well to the use of critical thinking. We saw some indication of this in language arts, particularly in the study of literature, where instruction is often less content-oriented and more focused on analyzing structures, inferring, comparing, and evaluating (although among these teachers, comparison was woefully lacking). The study of science and social studies should also require these kinds of thinking and indeed can provide very appropriate material to be the subject of such inquiries. But only math seems to do so consistently.

The discrepancy in math between written and oral assessment. The fact that inference questioning decreased so markedly and recall- and analysis-level questioning increased during oral classroom interactions is worth noting. The explanation for these large differences between written and oral assessments may enlighten our thinking about the teaching and assessment of higher order skills.

Here is a case in which a good match between how one teaches and how one assesses may not be totally necessary, or even desirable. Let's assume that in math class the teacher walks the students through the steps to solve the problem, asking questions along the way. Many of these questions solicit recall of facts and analysis of procedures, necessary for the ultimate

solution. In written assessment documents, on the other hand, the math teacher may often simply present the problem and expect a solution. The levels of questions posed in written assessments, therefore, are of a different order. Obviously, this could also occur in the teaching of other subjects. However, we have no other evidence of it in our data.

The lack of training in assessment. The teachers who participated in this study clearly lacked training in the assessment of higher order thinking skills. Not only did they tell us so, but it was also quite evident in the data descriptive of their assessments.

These teachers believe in the importance of teaching their students to think critically. For the most part they believe this can be taught. They are a little less certain about the assessability of such thinking skills. They candidly rate their ability to assess thinking skills as lower than their ability to teach them, but they are willing to do both. Other research we have conducted has shown that teachers' confidence and attitudes regarding assessing higher order thinking skills change dramatically with just one workshop on how to measure these skills (Stiggins 1987). The study we have reported here shows a clear need for more teacher training in this area.

The Utility of This Methodology

We designed this study to obtain a more accurate view of classroom assessment of thinking skills than has been obtained in prior research in this field. We have taken a multi-method approach which has included teacher interviews, classroom observations, and analyses of written assessment documents and their accompanying texts. With this approach we have succeeded in gaining a better understanding of the extent to which teachers, across grades and subjects, assess higher order thinking skills.

A particularly critical and unique part of our methodology is the observation and coding of oral questions as they are asked in the classroom. In the past it has been difficult to examine the reality of classroom assessment. Both the reliability and the validity of our methodology were carefully tested in the field test. We are confident in the capability of this methodology to classify questions accurately according to cognitive level. It is a unique and useful way to quantify what appears to be very qualitative data. By analyzing both the written assessment documents teachers use and the types of questions they ask during class, we have been able to document the need to improve the quality of classroom assessment of thinking skills.

This particular study has limitations, of course. Due to limitations in resources, we were only able to take a "snapshot" of each classroom--a day in the life of each teacher. We also had a limited number of teachers for each grade level and subject. Still, the utility of our methodology in determining the need for improved and expanded higher order thinking skills assessment suggests its potential for use in additional studies. If a number of school districts and a larger number of teachers were involved in each of the grade and subject categories, more generalizable patterns might be identified regarding the assessment of thinking skills at specific grade levels and within certain subjects. Schools could determine whether and in what areas their teachers needed further training. As this study has clearly shown, this is an area in which teachers do not yet tap the full range of their students' potential.

References

Baron, J.R., & Sternberg, R.J. (1985). Teaching thinking skills: Theory and practice. New York, NY: Freeman.

Carter, K. (1984). Do teachers understand the principles for writing tests? Journal of Teacher Education, 35(6), 57-60.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows to the classroom. In W.E. Hathaway (Ed.), Testing in the schools. New Direction for Testing and Measurement, No. 19. San Francisco, CA: Jossey-Bass.

Kearney, C.P., Kean, M.H., Raeber, E.D., Stevens, B.L., Baron, J.B., Fremer, J., & Daniel, M. (1985). Assessing higher order thinking skills. TEM Report 90, ERIC Tests Measurements Evaluation. Princeton, NJ: Educational Testing Service.

Natriello, G. (1987). The impact of evaluation processes on students. Educational Psychologist, 22(2), 155-175.

Quellmalz, E.S. (1985). Developing reasoning skills. A chapter in Baron, J.R., & Sternberg, R.J. Teaching thinking skills: Theory and practice. New York, NY: Freeman.

Schafer, M.D., & Lissts, R.W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38(3), 57-63.

Stiggins, R.J. (1987). The impact of teacher training on classroom assessment. Portland, OR: Northwest Regional Educational Laboratory.

Stiggins, R.J., & Bridgeford, N.J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4), 271-286.

Stiggins, R.J., Conklin, N.F., & Bridgeford, N.J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice, 5(2), 5-17.

Appendices

PRE OBSERVATION INTERVIEW RECORD

Teacher _____
Interviewer _____
Grade _____
Date _____
Subject _____

	1	2	3	4	5	6	7
BACKGROUND							
TIME SPAN							
COURSE TITLE/SUBJECT							
CLASS MEETINGS PER WEEK							
NUMBER OF STUDENTS							
GRADE							
NOTABLE CHARACTERISTICS:							
TEA. EST. OF ABILITY							
SPECIAL GROUPING							
SPECIAL LANG. ISSUES							
INSTRUCTION							
What is your goal for the period?							
What content will be covered?							
Is this new or review material?							
What text pages will be covered?							
What activities are planned? 1 = Bd wk 2 = Seat wk 3 = Whole Gp 4 = Small Gp							
ASSESSMENT							
Are any assessments planned? (note type)							
What kinds of questions will you pose? (note thinking skills)							

DOCUMENT ANALYSIS RECORD

BACKGROUND

TEACHER _____
ANALYST _____
GRADE LEVEL _____
COURSE/SUBJECT _____
DATE _____

DOCUMENT DESCRIPTION

Assessment Type:
___ Tea - Dev P&P Test
___ Text - Emb P&P Test
___ Written Assignments
___ Performance Assessment
___ Oral Questions
___ Standardized Achievement Test
___ Other (Specify _____)

Content: ___ M ___ S ___ SS ___ LA

Title _____

Author: ___ Teacher ___Pub. ___ Student ___ Other (Specify _____)

Use: ___ Classwork ___ Homework ___ Test ___Other (Specify _____)

Used on observation day? ___ Period ___

ASSESSMENT DESCRIPTION

HOTS ASSESSED:

	Items N	Items %	Test Points N	Test Points %
Recall	___	___	___	___
Analysis	___	___	___	___
Comparison	___	___	___	___
Inference	___	___	___	___
Evaluation	___	___	___	___
		100%		100%

RESPONSE MODE:

	Items N	Items %
Selection	___	___
Fillin	___	___
Essay	___	___
Product	___	___
Other	___	___ (Specify ___
		100% _____)

COMMENTS

RECORD OF PERFORMANCE ASSESSMENT

Teacher _____
Observer _____
Grade level _____
Subject _____
Period _____
Date _____

1. What skill was assessed?
   Process, describe _____
   _____
   _____
   _____
   Product, describe _____
   _____
   _____
   _____
   Combination, describe each above.

2. Were performance criteria:

Clearly definable by the teacher	___ ___ ___ ___ ___	Vague in the teacher's mind
Made very clear to the student	___ ___ ___ ___ ___	Not communicated to the student

3. List performance criteria (or attach list)

4. What thinking skills are reflected in the criteria? (Quantify if possible)

   Recall

   Analysis

   Comparison

   Inference

   Evaluation

Comments?

4777e

CODES

HOTS	Response	
R	T = Target	+ = Right
A	V = Volunteer	- = Wrong
C	G = Group	? = Unsure
I		± = Part right, part wrong
E		

ASSESSMENT INTERACTION CHART
(Page ____)

Teacher _____
Observer _____
Grade Level _____
Course/Study _____
Period _____
Date _____

START TIME	GROUP	QUESTION OR COMMENT	HOTS	RESPONSE

POST OBSERVATION INTERVIEW RECORD

TEACHER _____
INTERVIEWER _____
GRADE LEVEL _____
SUBJECT _____
DATE _____

ASSESSMENT PROCEDURES

Listed below are 6 types of assessment of student performance. After I define each, I will ask you to describe your backgound and use of that type of assessment.

COMMENTS

TEACHER - DEVELOPED PAPER & PENCIL TESTS & QUIZZES - M/C, T/F, MATCHING, FILLIN, & ESSAY TESTS & QUIZZES

		Math	Science	SS	LA
1. Use	Daily	___	___	___	___
	2-3 times/week	___	___	___	___
	Weekly	___	___	___	___
	Bi-Weekly	___	___	___	___
	Monthly	___	___	___	___
	Every term	___	___	___	___
	Annually	___	___	___	___
	Never	___	___	___	___

2. Importance	Unimportant	___ ___ ___ ___ ___	Very important
3. Confidence	Not confident	___ ___ ___ ___ ___	Very confident
4. Training	None	___ ___ ___ ___ ___	A great deal
5. Informed	Uninformed	___ ___ ___ ___ ___	Well informed

6. Thinking skills assessed (probe RACIE)

Math

Science

Social Studies

Language Arts

TEXT - EMBEDDED PAPER & PENCIL TESTS & QUIZZES - TESTS & QUIZZES THAT ACCOMPANY PUBLISHED TEXTS & MATERIALS

			Math	Science	SS	LA
1.	Use	Daily	____	____	____	____
		2-3 times/week	____	____	____	____
		Weekly	____	____	____	____
		Bi-Weekly	____	____	____	____
		Monthly	____	____	____	____
		Every term	____	____	____	____
		Annually	____	____	____	____
		Never	____	____	____	____

2. Importance — Unimportant ___ ___ ___ ___ ___ Very important

3. Confidence — Not confident ___ ___ ___ ___ ___ Very confident

4. Training — None ___ ___ ___ ___ ___ A great deal

5. Informed — Uninformed ___ ___ ___ ___ ___ Well informed

6. Thinking skills assessed (probe RACIE)

Math

Science

Social Studies

Language Arts

WRITTEN ASSIGNMENTS - Written work done by the student at home or during class time (including workbooks).

			Math	Science	SS	LA
1.	How often do you use this from of assessment to evaluate student achievement?	Daily	___	___	___	___
		2-3 times/week	___	___	___	___
		Weekly	___	___	___	___
		Bi-Weekly	___	___	___	___
		Monthly	___	___	___	___
		Once per term	___	___	___	___
		Annually	___	___	___	___
		Never	___	___	___	___

2. How <u>important</u> is this form of assessment in your scheme of student evaluation? Not important ___ ___ ___ ___ ___ Very important

3. How <u>confident</u> are you that these assessments accurately reflect student achievement? Not confident ___ ___ ___ ___ ___ Very confident

4. How much <u>training</u> have you had in the development and use of this form of assessment? No training ___ ___ ___ ___ ___ A great deal

5. How well <u>informed</u> are you about this form of assessment? Uninformed ___ ___ ___ ___ ___ Well informed

6. What kinds of questions do you ask? (probe RACIE)

Math

Science

Social Studies

Language Arts

PERFORMANCE ASSESSMENTS - MEASURES OF ACHIEVEMENT BASED ON BEHAVIOR OR PRODUCT OBSERVATIONS AND PROFESSIONAL JUDGEMENTS OR RATINGS

			Math	Science	SS	LA
1.	Use	Daily	____	____	____	____
		2-3 times/week	____	____	____	____
		Weekly	____	____	____	____
		Bi-Weekly	____	____	____	____
		Monthly	____	____	____	____
		Every term	____	____	____	____
		Annually	____	____	____	____
		Never	____	____	____	____

2. Importance — Unimportant ___ ___ ___ ___ ___ Very important

3. Confidence — Not confident ___ ___ ___ ___ ___ Very confident

4. Training — None ___ ___ ___ ___ ___ A great deal

5. Informed — Uninformed ___ ___ ___ ___ ___ Well informed

6. Thinking skills assessed (probe RACIE)

Math

Science

Social Studies

Language Arts

ORAL QUESTIONING - QUESTIONS POSED & ANSWERED DURING INSTRUCTION USED AS A FORM OF ASSESSMENT

			Math	Science	SS	LA
1.	Use	Daily	____	____	____	____
		2-3 times/week	____	____	____	____
		Weekly	____	____	____	____
		Bi-Weekly	____	____	____	____
		Monthly	____	____	____	____
		Every term	____	____	____	____
		Annually	____	____	____	____
		Never	____	____	____	____

2. Importance — Unimportant ___ ___ ___ ___ ___ Very important

3. Confidence — Not confident ___ ___ ___ ___ ___ Very confident

4. Training — None ___ ___ ___ ___ ___ A great deal

5. Informed — Uninformed ___ ___ ___ ___ ___ Well informed

6. Thinking skills assessed (probe RACIE)

Math

Science

Social Studies

Language Arts

7. How do you maintain records of student responses?

NORM - REFERENCED STANDARDIZED ACHIEVEMENT TESTS - PUBLISHED ACHIEVEMENT TEST BATTERIES

		Math	Science	SS	LA
1. Use	Daily	____	____	____	____
	2-3 times/week	____	____	____	____
	Weekly	____	____	____	____
	Bi-Weekly	____	____	____	____
	Monthly	____	____	____	____
	Every term	____	____	____	____
	Annually	____	____	____	____
	Never	____	____	____	____

2. Importance — Unimportant ___ ___ ___ ___ ___ Very important

3. Confidence — Not confident ___ ___ ___ ___ ___ Very confident

4. Training — None ___ ___ ___ ___ ___ A great deal

5. Informed — Uninformed ___ ___ ___ ___ ___ Well informed

6. Thinking skills assessed (probe RACIE)

Math

Science

Social Studies

Language Arts

THINKING SKILLS ASSESSMENT

1. How much training have you had in <u>teaching</u> higher order thinking skills, critical thinking skills or problem solving skills?

None ___ ___ ___ ___ ___ A great deal

Please describe that training

	HRS/TNG	DATE	FOCUS
Inservice training			
Graduate course			
Undergraduate course			
Personal study			
Other (Specify)			

2. How much training have you had in <u>assessing</u> higher order thinking skills, critical thinking skills or problem solving skills?

None ___ ___ ___ ___ ___ A great deal

Please describe that training

	HRS/TNG	DATE	FOCUS
Inservice training			
Graduate course			
Undergraduate course			
Personal study			
Other (Specify)			

3. What framework do you use in defining, teaching and assessing thinking skills?

4. What is the origin of this framework (i.e., where did you learn it)?

5. How do you feel about higher order thinking skills and their place in your classroom? Please think about these dimensions of your attitudes and rate them on the scales provided:

Meaning of higher order thinking skills (HOTS)	Unclear	___ ___ ___ ___ ___	Clear
Importance in student learning	Unimportant	___ ___ ___ ___ ___	Important
Teachability of thinking skills	Cannot be taught	___ ___ ___ ___ ___	Can be taught
Assessibility of thinking skills	Cannot be measured	___ ___ ___ ___ ___	Can be measured
Your ability to teach HOTS (given current skills)	Unable to teach	___ ___ ___ ___ ___	Able to teach
Your willingness to teach HOTS (given current skills)	Not willing	___ ___ ___ ___ ___	Willing
Your ability to assess HOTS (given current skills)	Not able	___ ___ ___ ___ ___	Able
Your willingness to assess HOTS	Not willing	___ ___ ___ ___ ___	Willing

6. Are you aware of district, building or department policies regarding assessment of thinking skills?

7. If so, do they influence your assessment? How?